NPS-AS-93-004

NAVAL POSTGRADUATE SCHOOL
Monterey, California

DATA MANAGEMENT:
Implementation and Lessons Learned from the Department of the Army Data Management Program

Tung Bui
Daniel R. Dolk
Christine M. Donohue
Gregrory A. Hayes

September 1992

Approved for public release; distribution is unlimited.

Prepared for: Director of Information, OASI (C3I),
Washington, D.C. 20301-3040

NAVAL POSTGRADUATE SCHOOL
Monterey, California

RADM R. W. West, Jr.
Superintendent

Harrison Shull
Provost

The report was prepared for the Director of Defense Information, OASI (C3I), Washington, D.C.. This research was funded by DoD Washington Headquarters Services, IAD, The Pentagon, Washington, D.C.

This report was prepared by:

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave Blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 1992	Technical Report, 1992

4. TITLE AND SUBTITLE
Data Management: Implementation and Lessons Learned from the Department of the Army Data Management Program

5. FUNDING NUMBERS
MIPR DXAM 20001

6. AUTHOR(S)
Tung X. Bui, Daniel R. Dolk, Christine M. Donohue and Gregrory A. Hayes

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(ES)
Administrative Sciences Department
Naval Postgraduate School
Monterey, Ca 93943

8. PERFORMING ORGANIZATION REPORT NUMBER
NPS-AS-93-004

9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Director of Information
OASI (C3I)
wASHINGTON, d.c. 20301-3040

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Data Management, Data Encyclopedia, Information Resource Management

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Data Management:
Implementation and Lessons Learned from the Department of the Army Data Management Program.

This study reports the lessons learned from the Department of the Army effort to identify and standardize data so that they can be shared across functional boundaries. Among other findings, this report concludes that data management is a continual process, and its success requires time and resources, and an effective communication between those who make the data management policy and those who implement that policy.

14. SUBJECT TERMS	15. NUMBER OF PAGES	16. PRICE CODE
	78	

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

CASE STUDY SERIES
ON IMPLEMENTATION PRACTICES OF
MANAGEMENT INFORMATION SYSTEMS
IN THE DEPARTMENT OF DEFENSE

Data Management: Implementation and Lessons Learned from the Department of the Army Data Management Program

Tung X. Bui
Daniel R. Dolk
Christine M. Donohue
Gregrory A. Hayes

Department of Administrative Sciences
Information Technology Management Curriculum
Monterey, California

September 1992

Acknowledgments

The authors would like thank Mr. Paul Strassmann, Director of Defense Information, OASD (C3I), for sponsoring the Case Study Series and the Department of the Army for their participation in this case study. Heartfelt thanks to Director of Information for Command, Control, Communications and Computers (DISC4), Deputy Chief of Staff for Personnel, and Army Information Systems Command personnel: Mr. Bruce Haberkamp, Ms. Cynthia Hanson, Ms. Chris Lundeen, MAJ Mike Napoliello, USA; Mr. Pete Pasek, and Mr. Russ Roberts for sharing their expertise. Special thanks to MAJ Bill Lasher, USA and Ms. Sheila Mudd of DISC4; Mr. Jim Glymph of the Army Information Systems Support Center, Data Management Directorate; and Ms. MaryJo Matera of the DoD Center for Data Administration for their unending patience, support, and encouragement.

Thanks as well to LT MaryJo Elliott, USN, for her assistance in the case study format; LTJG Bulent Bulbul, TUN, for his work on the figures presented throughout this case; Professor Sterling Sessions for sharing his interview techniques; and especially Professor James Emery for his painstaking review and insightful comments during the writing of this case.

Table of Contents

I. Executive Summary

Streamlining DoD through Information Technology

Sweeping changes are being felt throughout the Department of Defense (DoD). The end of the cold war has meant that the military services must redefine their missions and must be able to perform their new missions with fewer personnel and within the constraints of a seriously declining budget. One of the ways in which DoD can operate in this "leaner and meaner" environment is to leverage information technology.

DoD has embraced this philosophy through the Corporate Information Management (CIM) program. The focus of CIM is on enhancing the business process and improving the management of information. By understanding how DoD organizations operate — what processes they perform and what data they need to execute those processes — DoD managers will be able to streamline their business functions and develop information systems to meet the needs of the organization.

The Department of the Army (DOA) recognized nine years ago that data were a vital resource; it began formulating a policy to identify, standardize, and manage that resource. DOA began implementing its data management program in 1990. Although the Army data management program is in its infancy, the Army has started to realize the benefits of standardized data in the development of new information systems and have also been able to identify areas in which they can improve their business processes.

Movement Towards Data Management in the DOA

The Paperwork Reduction Act of 1980 mandated that executive agencies assign a senior information resource management (IRM) official and ensure that automatic data processing equipment is acquired and used in a manner that improves services and program management, increases productivity, and reduces waste and the information processing burden for the federal government. Federal agencies have been encouraged to develop strategic, tactical, operational, and information system plans to determine their information resource needs. The Army complied with these directives through Army Regulation (AR) 25-1, the Army Information Resource Management Program, but soon realized that successful planning for information resources depended on its ability to to identify and manage its data resource.

The Army began formulating a data management policy in 1984 while conducting a feasibility study for an Army corporate database. The implementation strategy for the corporate database called for the development of common data structures that would be defined in an Army-wide data dictionary. Although the corporate database project was canceled in 1988, the Army continued work on its strategy for a centralized approach to managing data and on the data dictionary as a tool to assist in implementing that strategy. The result was the Army Data Management and Standards Program (AR 25-9) and the Army Data Dictionary/Automated Dictionary Support System (ADD/ADSS).

The crux of the Army's data management program is to identify and to standardize data so that they can be shared across functional boundaries. The ADD/ADSS provides a mechanism to capture, standardize, and merge the information the Army needs to manage its data resources effectively. The ADD/ADSS was so successful that it was awarded a "golden nugget" by the Director of Defense Information and is currently being modified for use DoD wide.

Data Management Methodology in DOA

Using strategic data planning, the Army has been able to identify Army-wide data requirements and use those requirements as a basis for developing information systems that will meet its needs in the future. The Army's strategic data planning process is an iterative approach that uses modeling to represent Army business functions and information requirements. These models are initially developed at a high level to identify major business functions and essential information. Specific details regarding each business function and the information that is created or used in that function are captured in subsequent modeling efforts. The Army uses these models to develop blueprints for future information systems predicated on stable data needs.

Lessons Learned

The Army's experience in developing and implementing its data management program provides valuable lessons for other organizations within DoD:

- Management commitment is necessary for the success of any data management program of this scope.
- The data manager must be in a position of authority for the program to succeed.
- Effective communication between those who make policy and those who implement policy is a key ingredient to its success and effectiveness.
- The data manager, data administrator, and the database administrator must have the technical tools available to implement the data management policy.
- Modeling the organization and its data should be undertaken before standardizing data to determine what data are essential to the organization and who uses it.
- End-users play a critical role in the success of the data management program. They are the resident experts on how the business operates and what data are required to perform the job.
- An effective data management program takes time and resources. A sound data management program provides long-term benefits, but requires an up-front contribution of personnel, time, and budgetary resources.

- Data management and IRM is a continual process. Models and architectures must be updated as processes are re-engineered, new information systems are developed, or new technology is acquired if the organization is to gain long-term benefits.

Purpose of the Case Study

The purpose of the case study that follows is to:

- Acquaint the reader with the challenges and issues facing information systems executives
- Discuss the basic concepts of data management and key steps in implementing a data management program
- Present the Army's data management program and its efforts thus far in implementing the program
- Highlight the key lessons learned by the Army in implementing their strategy.

II. Management Information Systems in the 90's: Issues and Challenges

As information processing becomes more and more prominent in organizations, it is vital that computer systems be designed and implemented in a short time, without excessive costs, and meet the needs of the organization and its end-users. This is born out by a recent survey of information system (IS) executives, which lists the 14 top management issues they face for 1992.[1] These issues, listed in Table 1, can be aggregated into four main categories:

- Using information technology (IT) and data as a strategic resource
- Creating an information architecture that reflects organizational goals and objectives
- Integrating information systems
- Instituting Total Quality Management in IS

Industry is not alone in trying to cope with the problems related to IS. A recent survey details major and specific issues of paramount importance to the Department of Defense (DoD).[2] These issues, listed in Table 2, include:

- Strategic use of IS and data
- Integration of IS
- Information security and control
- Aligning IS with the objectives of the entire command

[1]Champy, 1992.

[2]Gacel, 1991.

- Aligning IS and Corporate Goals
- Re-engineering Business Processes through IT
- Creating an information Architecture
- Utilizing Data
- Improving the IS Human Resource
- Instituting Cross-Functional Information Systems
- Improving Software Development Quality
- Improving Leadership Skills in IS
- Boosting Software Development Productivity
- Developing an IS Strategic Plan
- Cutting IS Costs
- Instituting Total Quality Management in IS
- Integrating Information Systems
- Using IS for Competitive Breakthroughs
- Managing Dispersed Systems

Table 1. Top Management Issues of IS Executives for 1992

- Funding for IS

To bring focus to the above DoD issues, the Corporate Information Management (CIM)[3] initiative has become a driving factor in decisions regarding the development and use of information systems by DoD executives.

[3]For information regarding CIM see Brewin, 1991.

• Improving IS strategic planning
• Integrating data processing, office automation, and telecommunications
• Improving information security and control
• Making effective use of data as an organizational resource
• Aligning an IS activity with the objectives of the entire command
• Improving the quality of software development
• Facilitating and managing end user computing
• Increasing understanding of the role and contribution of IS
• Establishing a streamlined, more efficient procurement process
• Determining IS funding levels

Table 2. Top Ten Critical MIS Issues in DoD

Despite the recognition of critical issues by executives, the deployment of information systems has been and continues to pose complex problems to organizations. Many bottlenecks have been identified with respect to information systems development:

- Backlogs of several years for new systems development
- The cost of developing and maintaining systems
- Inability of management to obtain the information needed from computers to make decisions
- Redundant and often inconsistent data
- Timeliness and accuracy of data

In an attempt to deal with these pressing problems, concepts such as "re-engineering" and "information engineering" have been added to the portfolio of today's IS managers.

A. Data as a Strategic Resource

Regardless of whether the playing field is DoD or the private sector, two major themes stand out — 1) executives need timely and accurate information to make strategic decisions regarding the organization, and 2) IS must cross functional boundaries to provide access to that information.

As a strategic resource, information systems required to support the organization's mission need access to data that are increasingly distributed throughout the organization. For most organizations, data are hidden in applications and processed on many different platforms, making it difficult to obtain the integrated information needed for strategic decisions. Therefore, it is critical to adopt an Information Resource Management (IRM) policy that promotes an enterprise-wide view of data and uses data processing systems to put IS resources to work as strategic weapons.

IRM has evolved over several decades as organizations have learned how to assimilate information technology. IRM can be viewed as the policy arm for strategic systems. From a practical standpoint, however, IRM is still an elusive concept. We will define IRM as the process of directing or controlling the use of an information system comprising any combination of hardware, software, procedures, documents, or people that transforms data into a meaningful and useful form for satisfying organizational goals and objectives. It is important to realize that data themselves are a resource — in fact, the basic element from which information is derived.

Organizations must undergo a learning process to achieve maturity in their use of IT as a strategic resource. Ideally, an organization reaches maturity when its IT is fully integrated into the organization and when plans, policies, and control mechanisms reflect IRM concepts.

B. Evolution of IRM

IS researchers have proposed a number of models that try to capture the evolution of information technology within an organization.[4] These models provide a basis for information resource managers to ascertain where their organization lies with respect to computer technology and to development of realistic plans for IRM.

Common themes that appear in these models are organizational structure and the management of change in a technological environment. These themes lead to the following observations concerning the evolution of IRM:

- There is a correlation between organizational structure and the IRM strategy an organization should adopt for a given rate of technology diffusion. For a mechanistic organization such as the DoD, a reasonable IRM plan should ensure a low rate of growth during the initial stages of technological learning and then adopt a higher rate of growth once the organization begins to integrate technology.

- There are continual shifts between centralization and decentralization of IT. This shift is caused by the degree of control (financial, developmental, and operational) the organization exerts over the diffusion of technology and the rate of technological growth the organization desires.

- The growth of IRM follows organizational learning about computing technology. The organization moves through stages of technology assimilation[5], each stage building on the stage before it as the organization strives for equilibrium in its IRM strategy for that technology.

- Management activities within each stage should include policy setting, planning, support, and control for the IRM strategy and the current technology, as well as educating the organization, gaining upper management commitment, and ensuring end-user involvement.

- As a new technology is adopted, the organization begins the learning process again, with the goal of incorporating the new technology into the existing IRM philosophy. Thus there is no final stage for IRM, only maturity for an IRM strategy with respect to a certain technology. For example, organizations that have achieved maturity in the use of relational databases should start adopting an

---

[4]Conceptual frameworks such as the Stage Model developed by Richard Nolan, the Integrative Framework for End-User Computing (EUC) developed by Alavi et al., and Brown and Bostrom's model of EUC Management Effectiveness are excellent examples.

[5]Nolan identifies six stages of data processing growth as: Initiation, Contagion, Control, Integration, Data administration, and Maturity.

IRM strategy for Object Oriented Database (OODB), using what they have learned from the growth of relational databases and the IRM policies already in place.

- IR managers must have the technical tools available (data dictionaries, data encyclopedias, and I-CASE environments) to incorporate the IRM strategy.

C. Information Engineering Tools

If IRM is the policy arm for strategic systems, then information engineering is the vehicle for implementing that policy. Information engineering is based on the assumption that a relatively stable group of data lies at the center of the organization's information processing needs.[6] Tools used in information engineering provide support for data and process planning, analysis, design, and construction. It is this support for that make information engineering tools so powerful for modeling the strategic needs of the enterprise.

The basic tool for information engineering is the data dictionary. The idea of a central repository to store the organization's standard data definitions and data models evolved from database management systems (DBMS). As more and more organizations began to treat data as a valuable resource, it became apparent that DBMS-specific dictionaries could not support an overall systems development process with the organization's data needs as its focus.

Information engineering has led to the development of central repositories that not only store the organization's data, but also are used in conjunction with design tools such as Computer-Aided Software Engineering (CASE). CASE tools are intended to be a powerful extension of the data dictionary in that they automate certain processes in the software development lifecycle. Unfortunately, most CASE tools on the market today are not able to support the full software development lifecycle.

---

[6]Goodhue et al., 1992.

Information engineering techniques require a central repository that incorporates the features of a dictionary and a CASE tool repository. This has led to the concept of an encyclopedia, which is an extended central repository that:

- Supports any vendor's software development tool.
- Associates the metadata[7] it contains with applications resident on many different platforms.
- Automatically updates the organization's data architectures and models during software development.
- Enforces the organization's data management policies relating to the identification, naming, and maintenance of strategic systems.

Currently there is no tool available on the market that truly fits the definition of an encyclopedia. However, once such a tool is available it will serve as the reference point for the development and implementation of all information processes for the organization.

---

[7]Metadata is defined as data about the structure of data stored in the data dictionary. It includes such characteristics as the data name, format, and domain.

III. Data Management: Concepts, Methodology, and Issues

The basic source of information are data. The management of data must be the first step in developing an information systems architecture, making effective use of the data resource, and improving IS strategic planning. Data management is an integral part of IRM, which involves locating, organizing, cataloging, storing, retrieving, and maintaining data fundamental to the organization. This process consists of building models that reflect business functions and their associated information requirements. These models are not static, but change as the organization changes.

The purpose of this chapter is to provide the reader with basic concepts and issues in data management. A reader familiar with these may wish to skip this chapter.

A. Evolution Toward Data Management

The task of data management is to identify and control information that has strategic importance within an organization. The need for data control and management can be evidenced by the way organizations have developed information systems. This section traces the evolution of data management from file processing to corporate databases.

1. File Processing Systems

When a user or computer application requests data, the information system must know what data are stored, how they are organized, and how to access them.[8] File processing systems, shown in Figure 1, keep separate files for each application and the file layout is part of the application program.

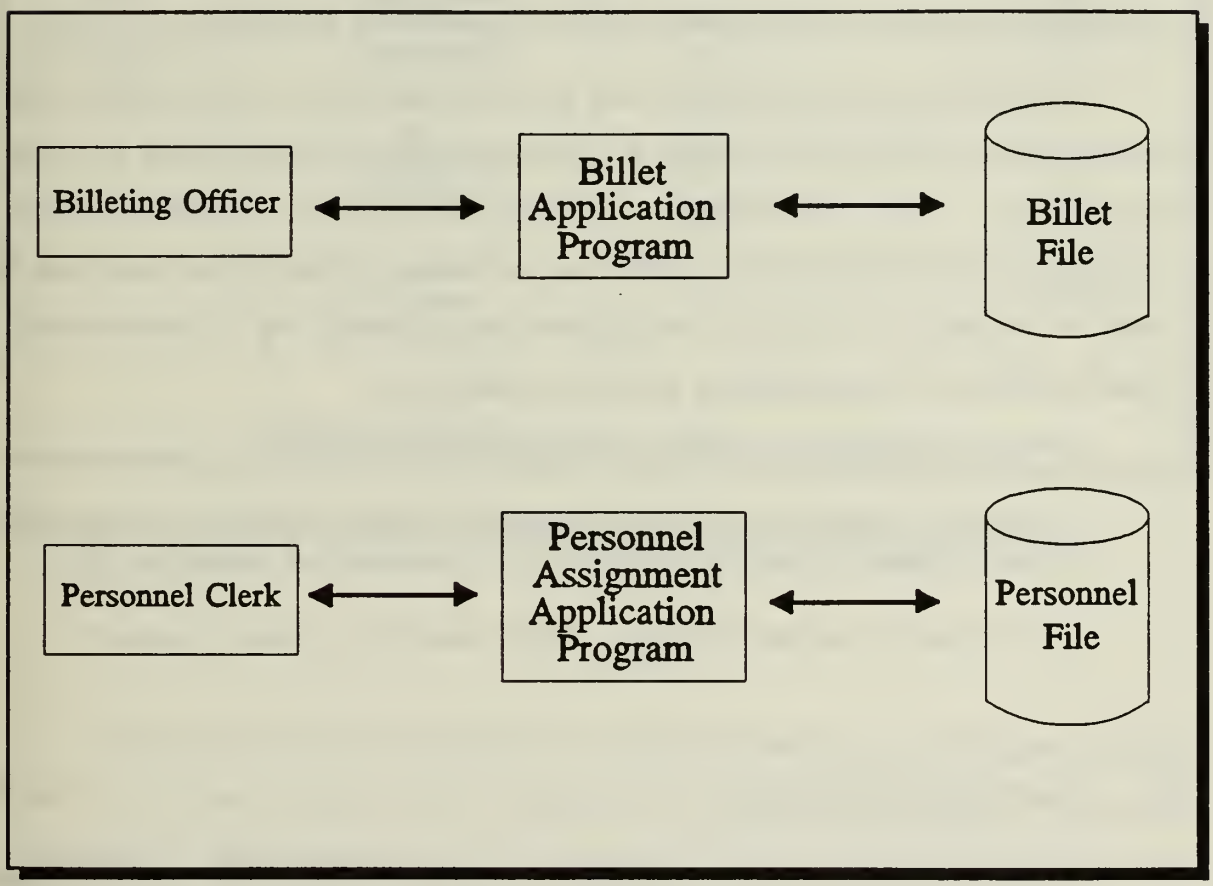

Figure 1. An Example of File Processing

In file processing systems, metadata about corporate information are embedded within the application program's source code. The only way to gain access to the

[8]Narayan, 1988

information is to access the application. For example, if the Billet application in Figure 1 were a COBOL program, it would contain a file section within the Data Division specifying basic information about each data element in the Billet file, including the size, the relative position within the record and file, and the access methods. For the Personnel Assignment application to make use of the information contained in the Billet file, it would have to know the data structure. Since this information is embedded within the Billet application, the Personnel application would have to access the Billet application file section. (or duplicate it within the personnel application).

To circumvent these problems, data are often duplicated in separate files. While this duplication may be efficient from a processing standpoint, the tradeoff is a loss of data integrity. Data have integrity only if they are consistent. Duplication leads to inconsistencies in the organization's data resource because updates to the same data that reside in separate files are often overlookedor are processed out of synchronization. Other limitations of file processing systems include:

- Data are separated and isolated, making integration difficult
- Application programs are dependent on file formats, therefore a change in file format requires a change to the application programs that access that file
- It is difficult to represent complex objects using file processing systems[9]

2. Functional Databases

As organizations learn that information needs to be shared, they realize that data must be independent of the applications that process them. Database Management Systems (DBMS) provide organizations with the technology for storing and retrieving data in a central and shared location. The most common DBMS are hierarchical, network, and relational. A DBMS is a set of programs that are used to define, process, and administer the database and its applications. Typical components of the DBMS are shown in Figure 2.

---

[9]Kroenke and Dolan, 1988.

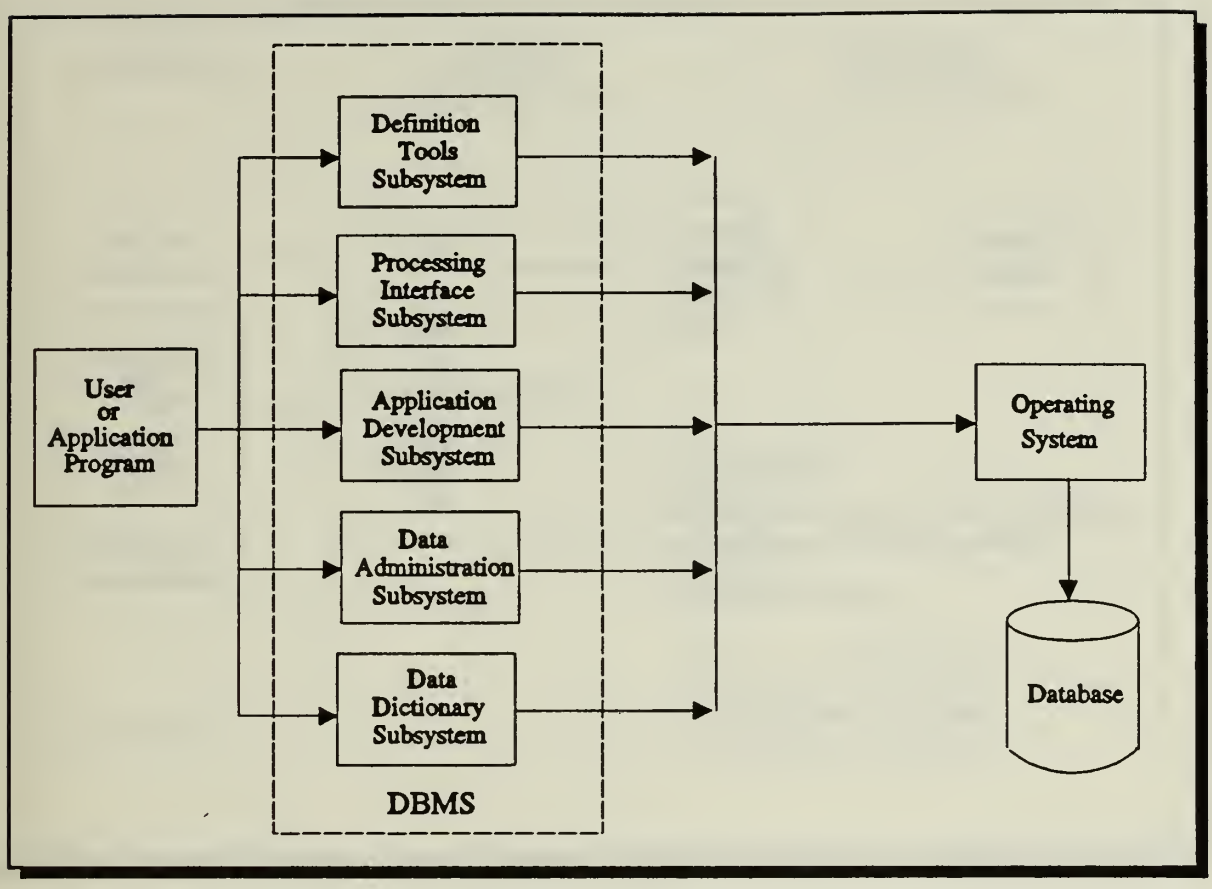

Figure 2. Typical Components of a DBMS (adapted from Kroenke and Dolan)

The decision as to what data are to be stored in the functional database is usually based on the data requirements of departments within the organization. Thus each department or functional area (e.g., personnel, finance, operations) develops and maintains its own database (See Figure 3). The functional area defines the structure of the database, called the schema, and develops a family of applications to process portions of the data, called subschemas. The subschema can be displayed to users on-line or in the form of reports so as to obtain meaningful information.

To define the database structure using the Definition Tools subsystem, entities[10] are defined in the functional work environment. Once the entities are identified, their

[10]An entity is any person, place, or thing that has meaning to a user.

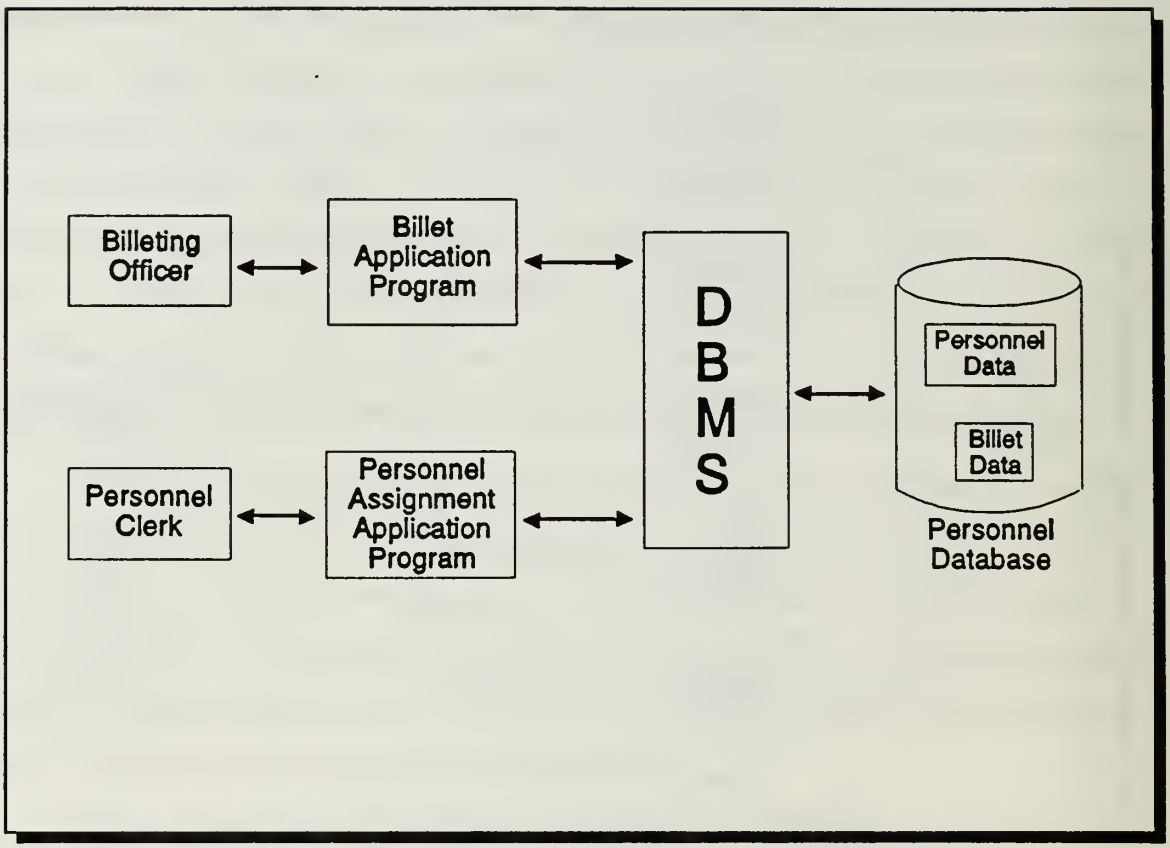

Figure 3. An Example of a Functional Database

attributes (characteristics) are specified. The attributes are further defined by specifying their domain[11]. For example, entities in the personnel department may consist of service member, service record, and dependents. The service member entity would be further defined as shown in Figure 4.

When all the entities are identified, an entity-relationship diagram is created to depict the relationships between entities. The list of entities and their attributes and domains, as well as the entity-relationship diagrams, comprise the database schema.

[11]Domain definitions specify formats, lengths, and special restrictions on the values of each domain.

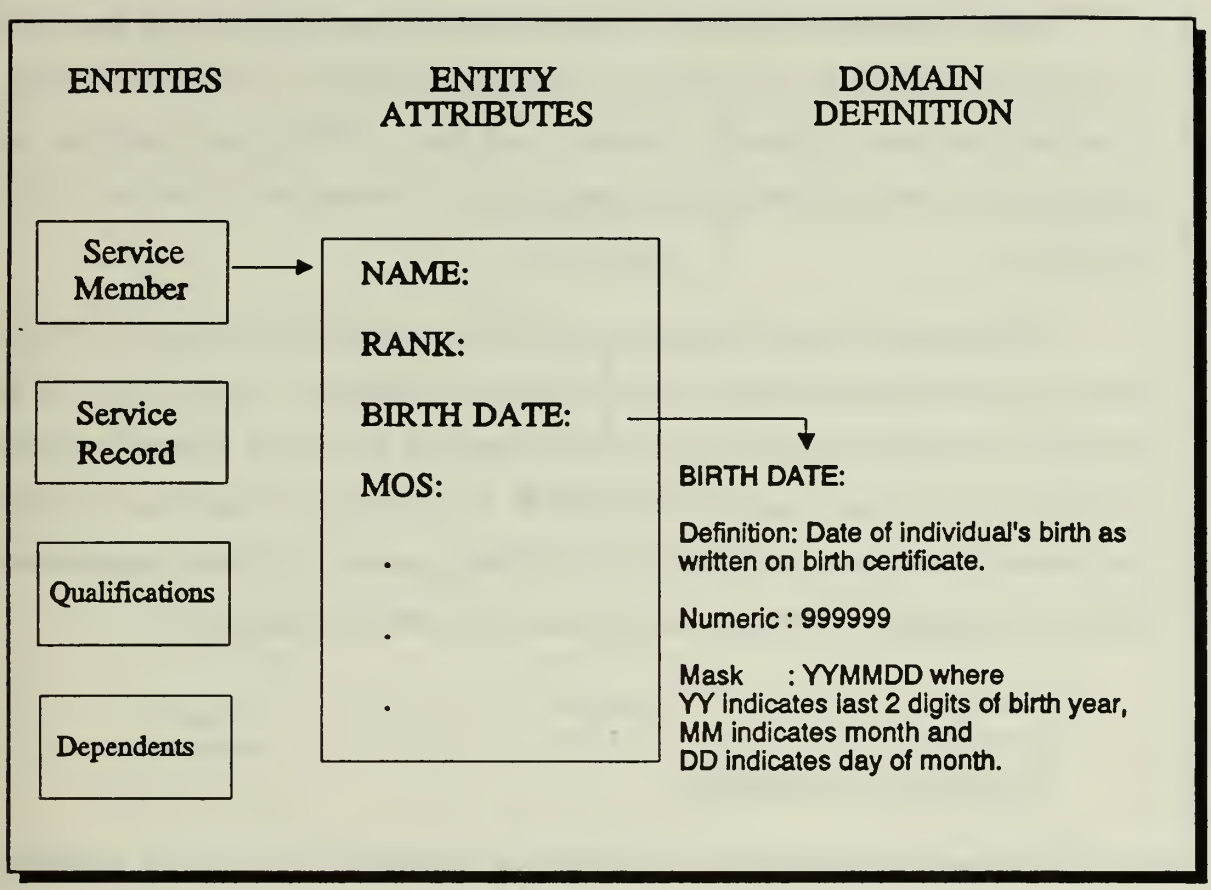

Figure 4. An Example of Entity, Attributes, and Domain Definitions

The data dictionary subsystem is a tool that facilitates storage of the organization's metadata and has the capability to relate that information in such a way that the organized metadata serves a useful purpose.[12] The data dictionary contains different types of information about data stored in the database. This information includes:

- Entity names and descriptions
- Data item names, descriptions, origin, format, and access rights
- Relationships between data items, entities, and application programs

[12]Narayan, 1988.

With the database technology described above, organizations have been able to use computer systems to represent their business environment, while minimizing the dependence between application programs and data. Within each functional area, information systems are developed so that data can be shared and duplication of data is minimized.

Unfortunately, many organizations that utilize database technology still do not use data as a global resource because accessing the data is difficult. Often, functional areas within the organization develop and maintain databases that follow different rules about how data are stored and accessed. In addition, the lack of an overall data management and standardization policy usually results in the organization's data being represented by different structures and different names within each functional database.

3. Corporate Databases

The need to share data across functional boundaries has led to the development of a central repository that should enable an organization to:

- Develop cross-functional applications independent of the data
- Promote the sharing of corporate-wide data through the use of common data structures
- Allow for access of data by heterogeneous databases

A central repository subsumes the features of a data dictionary as well as information about where and how data are stored in the databases. Figure 5 depicts the relationship of the repository to the organization's data.

Integrated information computer architectures, such as the Information Resource Dictionary System (IRDS) federal standard, provide facilities for recording, storing, and processing descriptions of an organizations's data and processing resources.[13] The use of such a tool enables data to be distributed and accessed from a heterogeneous network

[13]Schwartz, 1991.

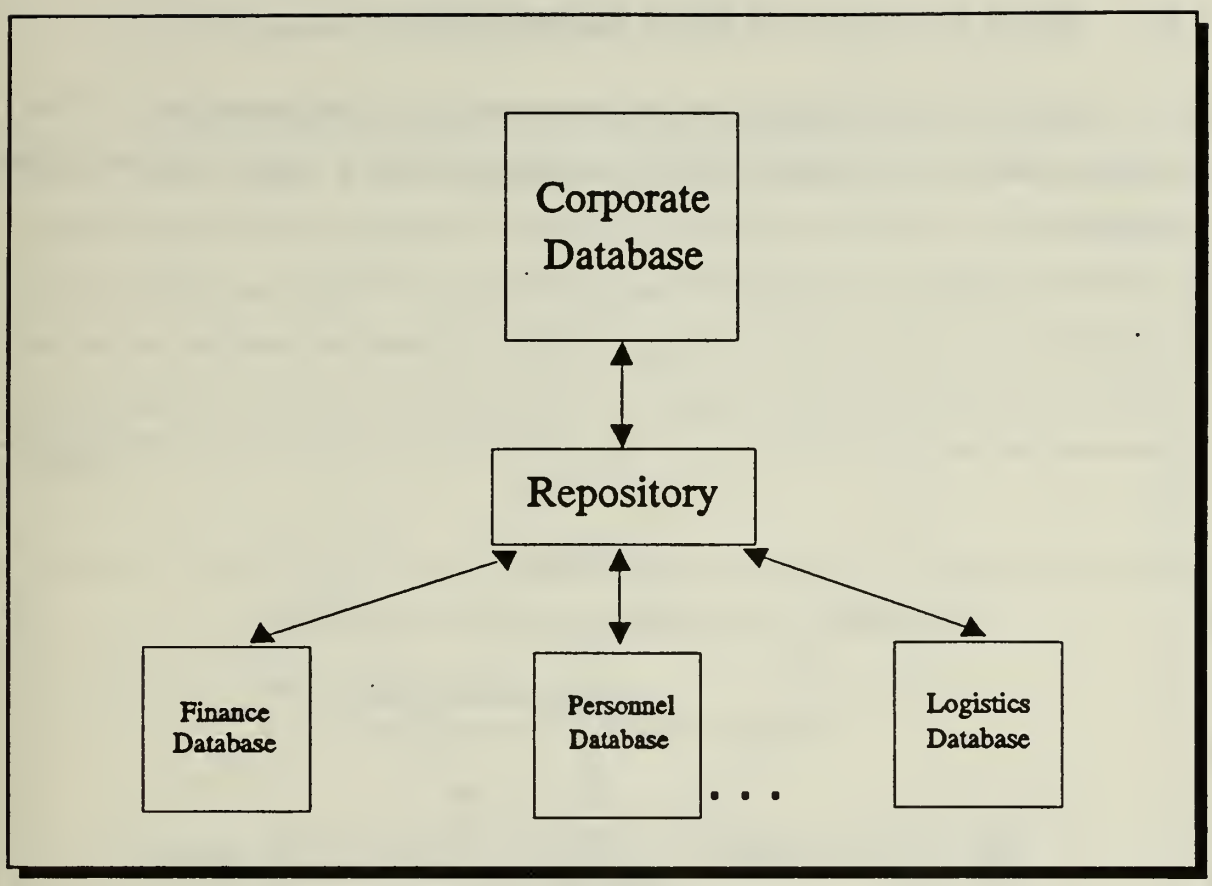

Figure 5. Repository and Database Relationship

of systems. In particular, the IRDS allows managers to:

- Develop and maintain corporate-wide data models that define entities, entity relationships, and process models, which will ensure the same entity definitions are being used throughout the corporation.

- Provide a central repository of information, a portion of which can be downloaded by application development tools such as CASE for the development of specific applications.

- Provide a means for using strict data-administration control procedures to ensure that changes to the local data model required by a new application will be incorporated back into the central data model.

- Coordinate database access and database access standards to ensure data security and integrity.

B. Basic Principles of Data Management

This section briefly discusses the basic concepts of data management. These concepts, depicted in Figure 6, form the foundation of a sound data management program.

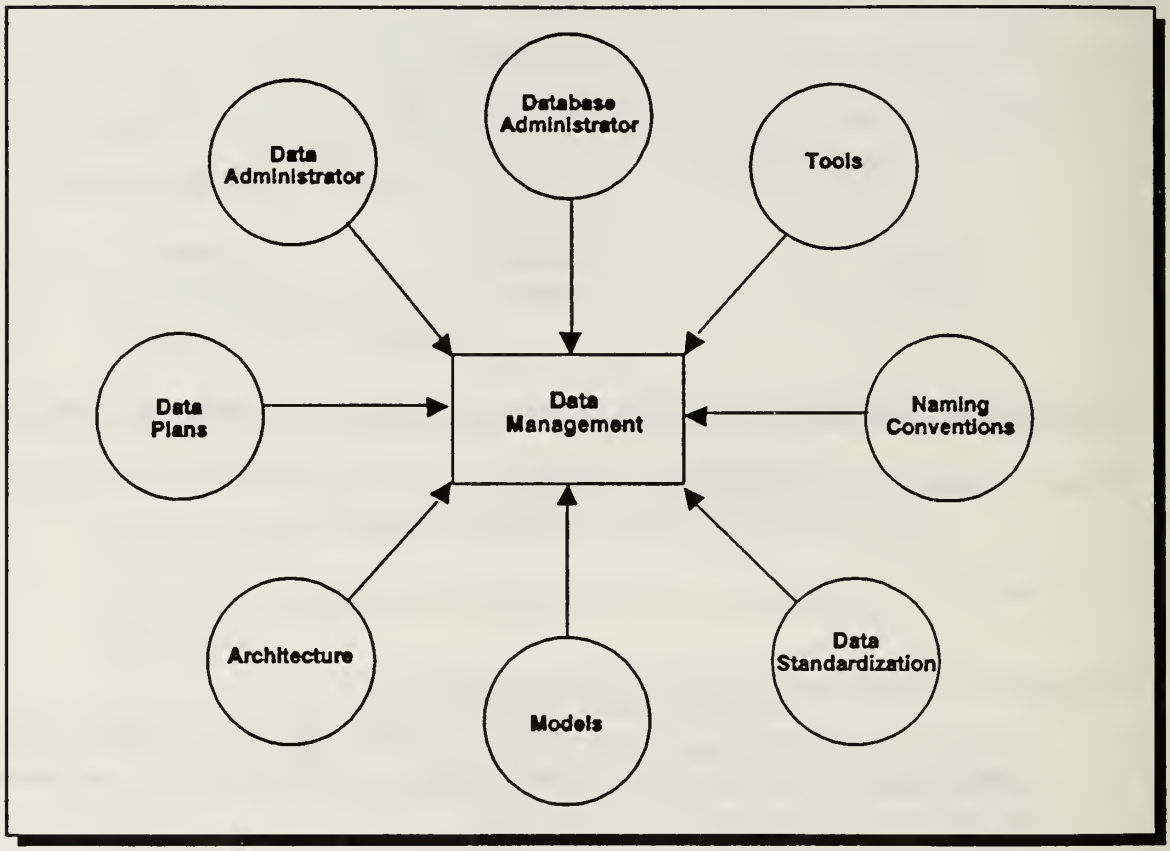

Figure 6. Building Blocks of Data Management

1. Key Personnel in Data Management

An environment of shared data requires a data administrator (DA) who has the overall responsibility for the organization's data resources. The DA implements methodologies for the centralized management and control of data resources. DAs and their assistants are responsible for planning and defining the conceptual framework for the overall database environment. The DA interacts with end-users to assess their data requirements in terms of the organization's needs. Functions of the DA are listed in Table 3.

Manage and Control Data Resources

- Strategic data planning
- Standardized data naming convention

Plan and Define Organization's Database Environment

- Functional and data architectures
- Functional and data models
- Design database structure

Provide Assistance to Database Users

- Training
- Support

Maintain Database Integrity and Availability

- Create database policy and planning procedures, documentation, and standards

Table 3. Functions of the Data Administrator

Database Design

- Evaluate technical needs
- Propose technical standards, design rules, and conventions

Database Implementation

- Database loading, testing, and validation
- Implement data definitions
- Implement and maintain data dictionary/encyclopedia and other database support software

Database Security and Integrity

- Install and maintain tools to guard against unauthorized access and unauthorized update, copying, removal, or destruction of the database
- Install and maintain tools to ensure the correctness and accuracy of data

Database Performance Monitoring and Evaluation

- Review, test and evaluate the performance of activity against physical data structures
- Initiate system improvements
- Assesses the impact of change
- Recommend database redefinition, redesign, and restructuring when indicated
- Implement restart, recovery, and backup procedures

Table 4. Functions of the Database Administrator

Where the DA requires skills in management, the database administrator (DBA) is the organization's technical expert on database related activities and has the responsibility for the operation of the organization's databases. Functions of the DBA are listed in Table 4.

2. Strategic Data Planning

Unlike the bottom-up approach of file processing and functional databases where data are identified because of the need for a particular application or business function, strategic data planning (SDP) is a top-down strategy for identifying and understanding *data* in the context of the overall business functions. This approach relies on modeling the organization, its processes, and its data as a basis for identifying and implementing an integrated set of information systems. The underlying assumption is that it is impossible to identify and develop information systems that will meet the needs of the organization without first knowing what the business is, what it does, and what data it uses.[14]

3. Functional and Data Modeling

To gain an understanding of the organization's processes and data, SDP approaches rely heavily on *modeling*. Two types of models are used in SDP. Functional models depict the activities or processes an organization performs to accomplish its mission. Data models describe the entities, their attributes, and interrelations that are necessary to support the business processes. Modeling is an iterative process with each session adding more detail until the model accurately reflects the processes and data associated with the business unit being modeled.

4. Information Systems Architectures

An architecture is a framework for specifying how components of a system fit together. An information systems architecture provides a framework within which future IS development, procurement, and implementation activities can occur. IS architecture

---

[14]Goodhue et al., 1988.

involves logical and physical elements.[15] Logical elements include the rules, procedures, and principles by which an organization operates. Physical elements of the IS architecture include applications, data/information, hardware/software processing platforms, and communications.

An applications architecture accommodates all activities within the organization, from operations to strategic planning. The data architecture is the glue that binds the other architectures together. It represents the information the organization must keep track of to perform its mission. The hardware/software architecture provides the processing power to run applications that will generate and distribute corporate data. The communications architecture connects the above three architectures so that information can be shared.[16]

5. Data Standardization

The role of a standard is to ensure that people and systems can communicate with each other in a consistent fashion[17]. Data standardization ensures not only that end-users and systems developers can communicate, but also that applications operating on heterogeneous platforms can exchange data effectively. Additional benefits of data standardization include:

- A means to control, share, and manage data
- Cost reduction of managing data by eliminating duplication

Once the organization's data model is completed, the entities and the data elements used to describe those entities should be standardized. The standardization process includes:

- Providing precise definitions for data entities and elements

---

[15]Kanter and Miserendino, 1987.

[16]Kanter and Miserendino, 1987.

[17]Narayan, 1988.

- Selecting entity and element names based on the organization's naming convention
- Identifying data element characteristics (context, format, domain)
- Identifying aliases
- Assigning responsibilities for specifications, definition, changes, etc.
- Identifying how the data are obtained (what process creates the data) and where they are used (what processes access the data)

6. Naming Conventions

Naming conventions help to establish consistency of data throughout the organization. There are several naming conventions currently being used.[18] No matter what naming convention is used, the element should be named by *what* it is, not *how* it is used. For example, a social security number is used as an account number for pay purposes and also as a number to uniquely identify an individual. Should the data element be named pay-account-number or individual-social-security-number? Naming the data element what it is (individual-social-security-number) promotes application independence and more accurately conveys its proper meaning, which will facilitate communication.

7. The Data Dictionary and Encyclopedia

The data dictionary and encyclopedia are indispensable tools for data management and the data administrator. The data dictionary provides a means to document, organize, and control an organization's information resources. As a tool for standardization, it enforces the integrity of the organization's data.

Used in the system development process, the data dictionary ensures that data standards are communicated and incorporated into the file structures and programs that

---

[18]The National Institute of Standards and Technology provide guidance for data naming.

are being developed. Documenting information about systems, programs, and data files reduces costs of future enhancements and maintenance of the systems.

Combined with a data dictionary, the encyclopedia provides the added benefit of storing the organization's data models and architectures and documenting their relationships with the organization's standard elements. The encyclopedia gives the data administrator a tool to analyze how a change in the organization's business processes will effect its data resource. A summary of the advantages of a data dictionary and encyclopedia is listed in Table 5.

Data Dictionary
• Provides a repository for standardized data
• Documents the relationship between data and information systems
• Provides control over organizational data
• Ensures data integrity
• Aids in system development
Encyclopedia
• Develops and stores organizational models
• Documents the relationship between models and data
• Analyzes the effects of change on organizational functions and data
• Aids in system development

Table 5. Benefits of a Data Dictionary and Encyclopedia

C. Data Management Methodology

The literature contains a number of approaches for managing organizational data.[19] The selection of a particular methodology will depend on the goals and objectives the organization has established for its data management and IRM programs. No matter what methodology an organization selects, there are certain key steps required in a sound data management program. These key steps are depicted in Figure 7 and are briefly discussed below.

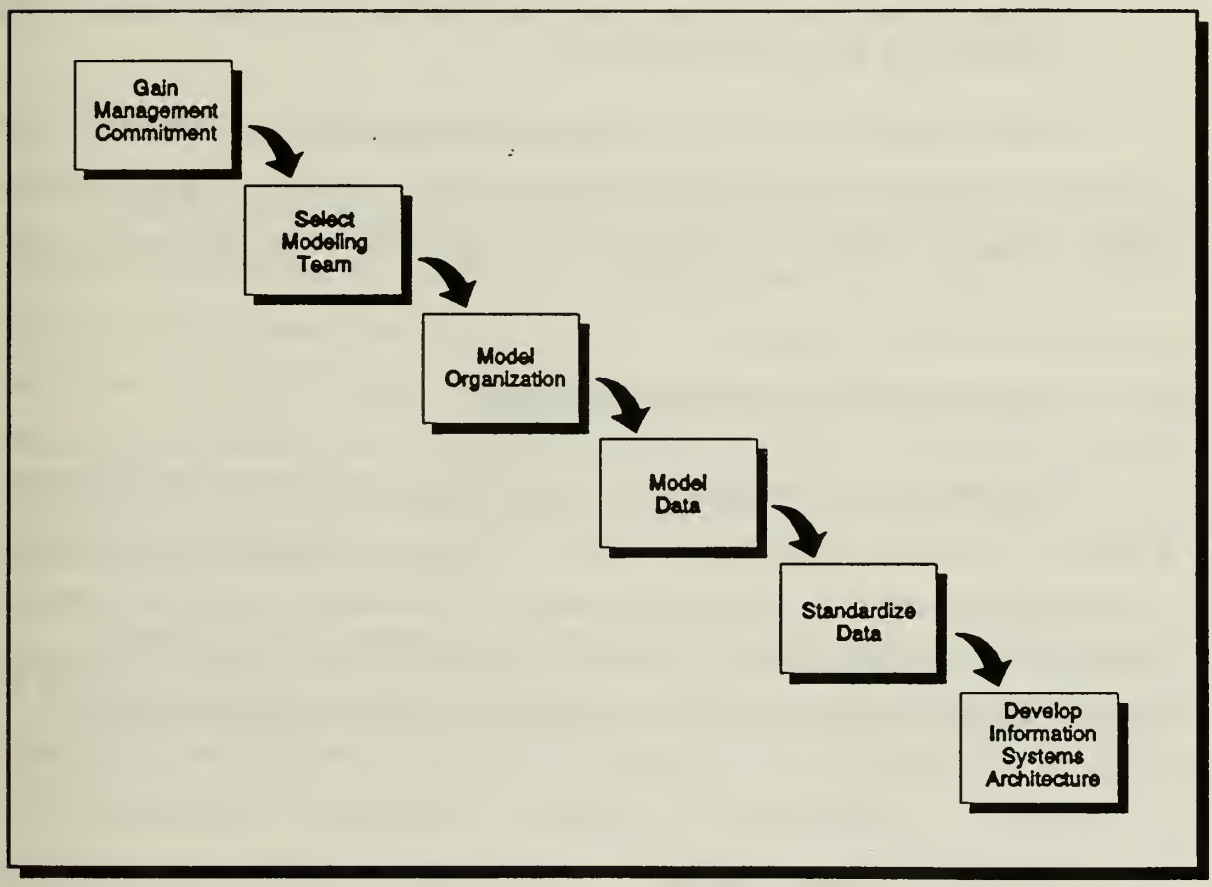

Figure 7. Key Steps in Data Management

[19]Examples include James Martin's Strategic Data Planning and IBM's Business Systems Planning (BSP).

1. Gain Management Commitment

Management commitment to a data management program is absolutely crucial. This commitment has to be given over an extended period without expecting immediate payoffs. Management commitment should be exhibited through:

- Dedication of resources — personnel and money
- Involvement and support of the strategic data planning process
- Support of standards, including the data modeling methodology and naming convention, as well as procedures for the control of changes to the program
- Support of the data dictionary and encyclopedia as the only source of data definitions for the entire organization[20]

Selling management may not be an easy task; however, the task can be made easier by conveying the fact that substantial time, labor, and money is spent duplicating efforts to create new files out of existing data and new applications that process the data.

2. Select Modeling Team Participants

Selecting the right participants is critical to the quality of the data management process. The modeling team should consist of at least one individual who is trained in the modeling process (a facilitator) and a group of functional personnel (end-users) who are experts in their particular area. The team may also include IS personnel as required, but care should be taken to ensure that IS personnel do not dominate the process.

3. Model the Organization

This step involves identifying business functions and the processes and activities within those functions that the organization performs. The major functions identified are grouped together in what is normally called an *enterprise model.* Each function in the

---

[20]Narayan, 1988.

enterprise model will be analyzed by performing functional modeling. During functional modeling, the business function is further broken down into processes and activities (both manual and automated).

The functional models will assist the organization in identifying processes that are currently supported by existing information systems, redundant processes and application programs that may be eliminated, and processes that may be candidates for information technology. In addition, these models allow the organization to look at its business processes in a fresh way and identify those processes that may benefit from re-engineering instead of, or before, automation.

4. Model the Organization's Data

With the functional model as a guide, entities used by the various processes and activities are identified and placed into a functional data model. The data model is further refined by associating the entities with the processes or activities that use or create them and by identifying relationships between entities. The data entities are also defined by documenting their definitions and characteristics. As each functional area completes its data model, it is integrated with other functional data models to identify data that can and should be shared.

The result of this integration is a corporate data model comprising entities that are of strategic importance to the organization. As with the functional data models, entity relationships are documented along with the processes that use the entity. In addition, the functional area is designated that will have responsibility for gathering and maintaining data on the individual entities.

5. Standardize the Data

In order for data to be shared, they must have a common structure and identification. Standardizing entities involves naming the entity using the organization's

naming convention, providing a precise definition, identifying its characteristics (attributes), and developing access keys. In many cases, a functional area will not require access to the entire entity, but rather a subset (called a *view*) of the entity's characteristics. These characteristics, referred to as *data elements,* must also be included in the standardization process.

6. Develop an Information Systems Architecture

When the functional and data models are complete, the organization will have a detailed picture of its business processes, how those processes gather, distribute, and use information, and what information is important to organizational goals and objectives. The modeling effort provides the organization with the opportunity to redesign the business based on information needs and build systems that will support the organization dynamically.[21]

The information systems architecture provides the organization with the framework for re-engineering the business. The IS architecture represents an organization's current and future applications, hardware and software, and communication and data needs. These architectures assist the organization in developing a set of plans for migrating to the future with respect to its basic business processes and its supporting information systems. In practice, it may be useful for the organization to sketch a preliminary information systems architecture to use as a road map during the first five steps of the data management process. However, we believe that a detailed information systems architecture cannot be completed without a thorough understanding of the organization's processes and its data.

---

[21]Finkelstein, 1991.

D. Issues in Implementing Data Management

The previous sections discussed the basic concepts of data management and the key steps in implementing a data management methodology. However, putting these concepts into practice still poses considerable problems for organizations.

1. Centralization versus Decentralization

The past three decades have seen continual tension over how information resources should be controlled. Management's desire to control information resources has always tended towards some degree of centralization; opposing this tendency, the proliferation of end-user technology, the need to access information in a timely fashion, and the size of organizations has placed pressures on management to decentralize resources. There are advantages to both strategies. While centralization appears to provide better security, integrity, and control, decentralization favors innovation and facilitates end-user support.

Total Quality Management (TQM) and re-engineering advocates support "decentralized centralization". This philosophy treats dispersed resources as if they were centralized. The use of a centralized database created by a sound data management policy enhances data security, integrity, and control. By applying technologies such as telecommunication networks and implementing standardized processing systems, the relevant portions of an organization's database can be downloaded into a functional database to provide end-users with the flexibility and service they require.[22] The issue is one of balance. Management must consider the needs of the organization with the needs of end-users to determine the extent of which the organization's information resources are decentralized.

---

[22]Hammer, 1990.

2. Data Standardization

Standardization invariably creates conflicts over ownership of data. If end-users do not believe data are a corporate resource, they will be unwilling to share those data with others. Information is power, and the person who owns the information holds the power. Along the same lines, end-users are reluctant to surrender ownership of data that they believe belong to their functional area for fear that the data will no longer be reliable. In some cases, no functional area will claim responsibility for the data, even though many functional areas use it frequently. Standardization requires that some group takes responsibility for the data element's lifecycle. Infighting can occur over which functional area should have stewardship over a data item.

Second, data element naming can also result in conflicts between end-users, especially when elements are shared by several functional areas. The idea that a data element will be called anything other than the name familiar to a particular user may also cause conflict with end-users. The data administrator must be prepared to address and handle these problems.

3. Data Security

Security of data and databases has always been a critical and complex issue, especially in DoD. Security for dictionaries and encyclopedias compounds this issue because the organization's metadata resides in one location. Unauthorized access to the encyclopedia and dictionary could result in compromise or a loss of data. In addition, although data element structures and relationships between entities and applications may be unclassified by themselves, the aggregate of that information may very well become classified. By having access to the metadata, users (authorized or unauthorized) could glean classified information. Dictionaries/encyclopedias constitute a single point of vulnerability and thus should be addressed in the organization's security plan.

4. Data Integration

Data integration across heterogeneous platforms can be achieved in three ways: interfaces, common access to data, and common control and access to data.[23] In interface-level integration, data are passed from one platform to another through an interface. If data are changing on both sides of the interface, maintaining consistency between platforms can be difficult.

Integration through common access to data greatly increases the level of coordination between products. Since different systems are unaware of the existence of other systems that use the same data, there is no mechanism in place to ensure that the changes introduced by one system are consistent with changes introduced by other systems.

To address the problem of multiple platforms sharing data, common control of data access is necessary. Controls must ensure that data being entered are consistent both with the metamodel and with data that already exist, and that changes to the data are regulated and tracked.

5. Data Synchronization

Synchronization refers to the consistency, accuracy, reliability, and timeliness of replicated data in distributed environments. With the sharing of data across distributed databases, users may be reluctant to make decisions from data over which they do not have control. Organizations will need to develop policies and procedures that ensure timely updates to data as well as synchronizing those updates when copies of the database are distributed.

---

[23]Aranow, 1989.

IV. Evolution of Data Management within DOA

The previous sections presented the issues and challenges faced by organizations in dealing with information systems and theoretical concepts that may be applied to meet those challenges. The Army has addressed some of the problems of managing information systems by developing a comprehensive program to manage data. The next two sections will present the Army's data management program and provide some valuable lessons other organizations may wish to consider before implementing a data management program of their own.

A. DOA Information Technology

Like any large organization, the Army maintains a considerable amount of data concerning its personnel, equipment, installations, and finances. The majority of the application software currently operational that processes data consists of "stovepipe"[24] systems. The Army is spending about $1.7 billion a year to maintain and operate its "sustaining base"[25] automation. This includes 3,700 applications amounting to almost 200 million lines of code, 96 percent of which is unique to specific commands and installations.[26]

[24]Systems developed to meet one functional area without taking into account other functional areas that potentially could use the same information.

[25]The Army defines Sustaining Base as: The area and information resources outside of the area of operations that have the responsibility to raise, organize, train, equip, and eventually deploy and sustain Army forces in the accomplishment of their mission in operational theaters.

[26]Rogers, 1991.

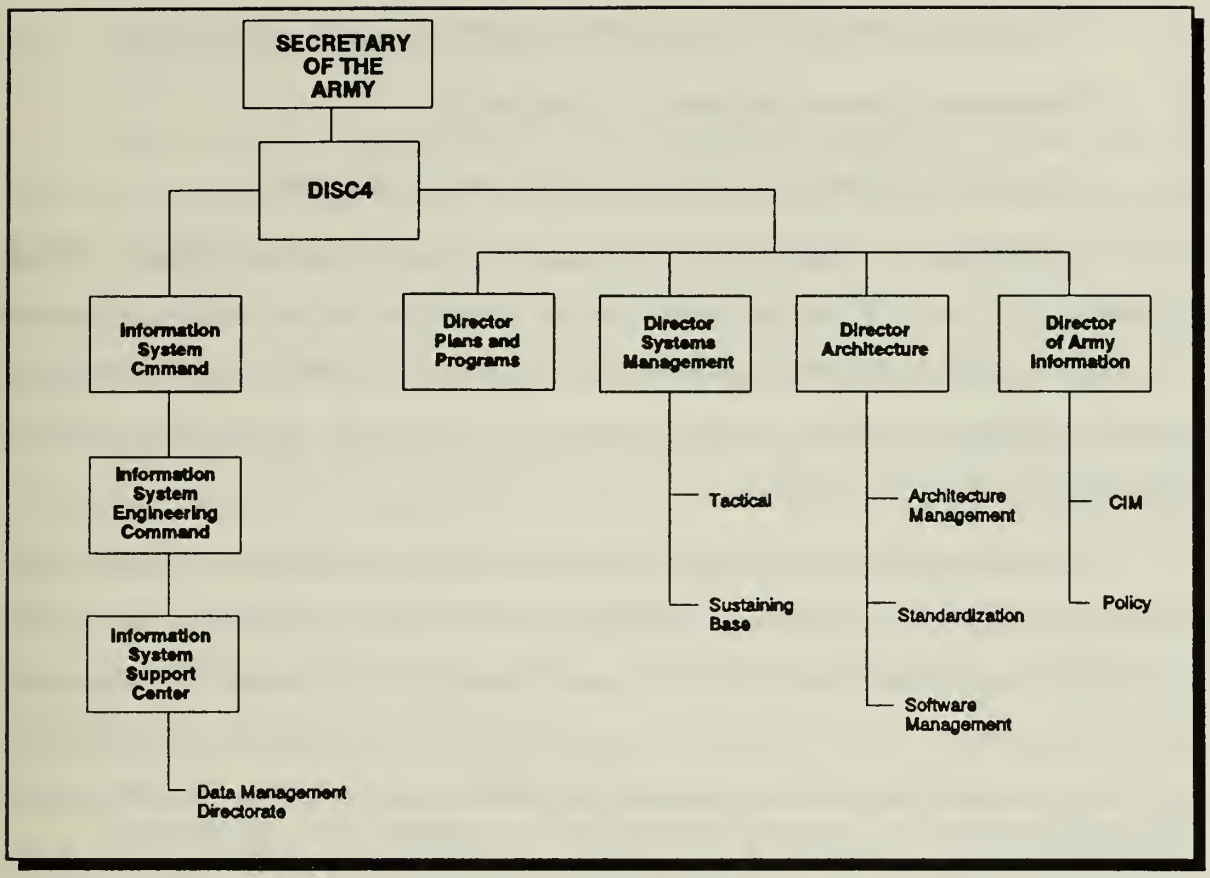

Figure 8. Army Key Information Management Commands

Army information resource management budget for Fiscal Year 1992 is $2.4 billion. IRM employees number approximately 17,000. Key Army Information Resource Management Program (AIRMP) commands are depicted in Figure 8.

The Office of Director Information Systems for Command, Control, Communications and Computers (ODISC4) is the policy arm for the Army's IRM and Data Management program. DISC4 is primarily responsible for:

- Information Management
- Plans and Policies
- Information Architecture and Models
- Information Requirements Development

- Functional Information Requirements Integration
- Information Systems Programmatic Integration
- Information Mission Area Program and Resource Integration

Subordinate to DISC4 is the Architecture Directorate and the Director of Army Information. The Architecture Directorate Standardization Division is responsible for the Army's Data Management and Standards Program (AR 25-9), while the Director of Army Information Policy division oversees the Army's Information Resource Management Program (AR 25-1).

The United States Army Information Systems Command (USAISC) is the Army's primary provider of information management staff assistance and information systems operational support and services in the Army Strategic Environment[27] and Sustaining Base Environment. This command is responsible for establishing and maintaining the technical interface between Army information systems within the Strategic and Sustaining Base environments, between environments, and with those external to the Army. USAISC is also responsible for technical information systems integration and the Information Mission Area (IMA)[28] systems engineers.

USAISC was appointed the Army Data Administrator by DISC4 and is given operational responsibility. This responsibility is delegated to the Data Management Directorate (DMD) of the Information Systems Software Center (ISSC), Information Systems Engineering Command (ISEC). As the Army Data Administrator, DMD is responsible for the Army-wide implementation oversight of the Army Data Management and Standards Program (ADMSP).

---

[27]Within information mission area (IMA), the environment that links the Theater/Tactical and Sustaining Base Environments.

[28]The resource requirements and associated information management activities employed in the development, use, integration, and management of information. Information resources include doctrine, policy, data, equipment, and applications, along with related personnel, services, facilities, and organizations.

B. DOA IRM Policy

Prior to 1984, management of Army information resources rested with the principal user of the information. Under this structure, little centralized control existed. Rather, individual users defined their particular information requirements, developed software, procured necessary hardware and communications, and linked these elements together to form information systems, usually fulfilling narrowly defined needs.[29]

The Paperwork Reduction Act of 1980 encouraged federal agencies to view information as a valuable organizational resource and to develop managerial approaches to improve its collection, storage, and dissemination[30]. In order to carry out the provisions of this act, the Army established The Army Information Resources Management Program (AR 25-1) in 1984. The 1988 version of AR 25-1 specifies the following objectives of the AIRMP:

- Establish a concept of operation and management processes and structures for the management of information and information resources that ensures integration, sharing, standardization, interfacing, interoperability, timeliness, and accuracy of information provided to Army decision makers in peace, transition to and from conflict, and conflict.

- Ensure that appropriate, timely, and accurate information is identified and made available for satisfying user requirements in the execution of Army and Army supported responsibilities.

- Ensure that existing information resources are identified, information requirements are validated, and a systematic approach for satisfying these requirements is established and maintained.

- Ensure that it is applied in the management of all IMA disciplines in all environments under all conditions for the Total Army.

To implement the AIRMP, the Army developed the Army Information Architecture (AIA). The AIA is the framework for developing information systems to support the Army's present and future mission requirements. The AIA structure

---

[29]GAO/IMTEC-90-58 "Army Information Resource Management"

[30]Head, 1990.

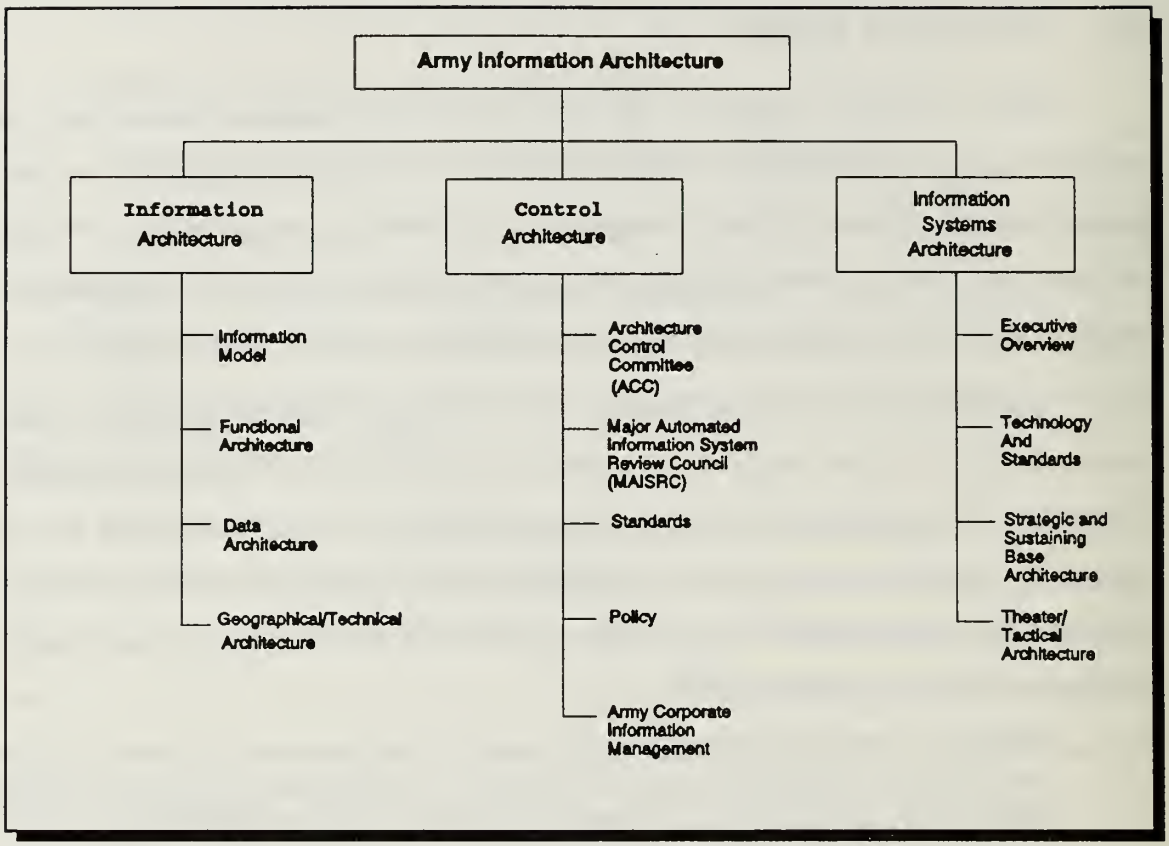

Figure 9. Army Information Architecture

(Figure 9) consists of three architectures: Information Architecture, Control Architecture, and Information Systems Architecture.

The total AIA is a composite of information architectures of major organizations[31] within the Army. Each organization is required to develop and maintain an organizational information architecture. Organizational information architectures reflect the requirements of the organization as well as subordinate activities and are guided by the information architecture of the next higher echelon.

[31]Major organizations include Headquarters Department of the Army agencies and their field operation and staff supporting activities; and Major Commands and their subordinate commands, installations, activities, and deployable units down to division and separate brigade level.

The Information Architecture is a blueprint for developing plans and actions in the planning, control, and management of information. It consists of an information model and a functional, data, and geographical/technical architecture.

The information model documents Army activities and information used or created by Army activities in support of Army missions and goals. The information model represents an organization's processes, information classes (ICs)[32], organizational units, and their relationships.

The functional architecture is a decomposition of the process groups listed on the information model. It is used as a framework for current applications management and future application development. The data architecture represents an analysis of the information model and is used as a framework for organizing information. The geographic/technical architecture is the framework for managing the geographic, communications, and technology implications of data and applications distribution.

The information systems architecture (ISA) is the physical implementation of the logical representation provided by the information architecture. This architecture provides technical guidance for modernizing IMA information systems. The ISA incorporates emerging standards, technologies, and mission changes to overcome deficiencies in the Army's information systems baseline configuration and to meet the IMA goal. The control architecture ensures the physical architecture (ISA) implements the logical architecture (i.e., the information architecture).

C. DOA Data Management

In 1983, the Vice Chief of Staff announced the need for a corporate data base, highlighting the adverse effects that conflicting and inconsistent data were having on Army-wide decisions. The corporate database project, although never completed, brought to the forefront the need for a data management program that would support the

---

[32]The Army defines an information class as a category of logically related information that supports the things of lasting interest about which the organization wishes to keep data.

Army IRM program through the use of common data structures throughout the Army. In September 1989, the Army established the Army Data Management and Standards Program (AR 25-9). This program implements data management for the IMA and the AIRMP. According to AR 25-9, the data management program will:

- Ensure that the mission-essential data requirements of commanders and decision makers are identified, documented, and supported.
- Exploit data as a shared resource to improve mission effectiveness and efficiency over the spectrum of conflict.
- Set the standards for accuracy, security, and synchronization of data in Army Data bases and information systems.
- Develop an understanding of data storage and access requirements that will be useful in developing standard Army information systems and data bases.

The Army believes that by identifying mission-essential data, exploiting data as a shared resource, and setting standards for data and Army information systems they will be able to:

- Manage data effectively and efficiently throughout their life cycle.
- Establish and maintain data architectures that support the Army's information requirements.
- Promote the development of applications independent of the data.
- Maintain and control data in databases so they are accessible by many applications.
- Promote the sharing of data by combining similar user data requirements and establishing interoperability requirements.
- Work to promote the smooth movement of shared data among the three IMA environments: strategic, theater/tactical, and sustaining base.

The Army data management program consists of six activities. These activities and their objectives are listed in Table 6. The Army has developed specific guidance for their strategic data planning and data elements standards. This guidance is provided in Army Data Management and Standards Program Administrative Procedures Guide (DA Pam 25-9-1). The strategic data planning and data elements standards activities, as

Activity	Purpose
Strategic Data Planning	Develop and maintain long-range planning documents which reflect data-related information requirements.
Data Element Standards	Develop policies and procedures for creating, standardizing, storing, and managing sharable Army data and data resources.
Information Management Control	Coordinate data requirements between the data management program and the Management Information Control System.
Data Security	Develop policies and procedures required to protect data and information. Data security includes the physical protection of data, control of user access and the collection and dissemination of information.
Data Synchronization	Develop policies and procedures that govern the consistency, accuracy, reliability, and timeliness of data used and generated by the Army. It also addresses the planning, storage, scheduling, and maintenance of data and the exchange of data among authorized users and systems.
Database Development and Maintenance	Develop policies and standards that guide design, development, documentation, and integration of databases. It includes standard procedures and methods for developing consistent database maintenance practices.

Table 6. Army Data Management Activities

outlined in DA Pam 25-9-1, will be discussed in greater detail in the following sections. The Army is still developing guidance regarding the last four activities.

D. Strategic Data Planning

Strategic data planning is the process that the Army has selected for their data management strategy. This planning process produces a plan that addresses:

- How Army organizations will develop and maintain a set of models and architectures that represent the organization and its information requirements.
- How the data-related information requirements of the organization will be met as it moves from its baseline configuration to its objective.

The plan also includes an initial set of functional and data architectures and models that represent the organization.

Strategic data planning functions are conducted by major Army organizations and occur during the four phases of the Army's information engineering process. These phases include:

- Information Requirements Study (IRS)
- Information Requirements Study Implementation (IRSI)
- Functional Area Analysis (FAA)
- Information System development

The IRS is a formal study conducted to determine organization-wide information requirements. During the IRS, high-level functional and data modeling are conducted and the initial organizational information model is developed. Further analysis of the organization's information model is conducted in the IRSI. In the IRSI phase, the functional and data models developed earlier are expanded and used to create a more detailed information model as well as the organization's initial functional and data architectures.

During FAA, a particular functional area is studied and modeled in significant depth. Detailed data and functional models are constructed. Data models developed during the FAA are integrated into the enterprise-wide data model. When the organization determines that there is a requirement for a new information system, models produced by the strategic data planning process are expanded to highly detailed data and

activity models for the system under development. Data requirements for the new system are compared with the enterprise and functional data models. Any data element that does not exist as a standard is developed using the Army's standardization procedures and added to the organization's information model.

1. Modeling Using IDEF

The focus of the Army's strategic data planning is on modeling — specifically, the development of a set of models representing the organization and its data. There are two types of models developed during the Army's strategic data planning process: functional models and data models.

The Army uses functional models, also know as "process" or "activity" models, to show a hierarchical breakdown of the activities undertaken by an organization. Army functional models identify inputs, outputs, controls, and mechanisms involved with each of the activities in the organization.

Data models are used by the Army to depict the entities of principal concern to the organization and the way the entities relate to each other. The data model is later used to help determine and define the data the organization must track, and provides the framework that enables the data to be standardized.

The methodology the Army has selected for modeling is IDEF[33]. IDEF is a methodology that maps out the primary activities of strategic data planning. This methodology consists of two parts:

- IDEF0: used for development of functional models.
- IDEF1X: used for development of data models.

---

[33]IDEF stands for ICAM (Integrated Computer Aided Manufacturing) Definition Language. It is beyond the scope of this paper to provide detailed documentation of the IDEF methodology. For information concerning the IDEF methodology see "Corporate Information Management Process Improvement Methodology for DoD Functional Managers".

2. Army Data Model

The Army Data Model is an integrated consolidation of all the functional area data models developed during the functional area analysis phase in information engineering. The Army data model serves as the basis for data definition, integration, and sharing across the Army and includes the business rules concerning Army data. Any system that creates or uses shared data will base these data on the Army data model. The Army data model is intended to be the Army's conceptual schema for corporate data. The conceptual schema represents a view of the data independent of both users (or applications) and systems (software and/or hardware).

E. Data Element Standards

One of the principal uses of data models is to serve as the basis for creating standardized data elements. Standard data elements are attributes of entities that have been defined and documented during the modeling process. The goals of Army data standardization are:

- To provide an Army-wide data framework for information system development
- Facilitate the sharing of data
- Support the integration of systems

The Army uses a four-step procedure to standardize data elements, as discussed below.

1. Data Identification

The need for a data element may be discovered during the modeling process of one of the information engineering phases or during an information system development effort. Once the requirement for a data element is identified, it is placed into the data model under the logical entity the data element describes. If the correct entity does not

exist in the model, the developer proposes an addition, expansion, or correction to the model to include the correct entity. Information Class Proponents must approve this change and the change must be fully integrated into the Army data model.

A search is conducted of existing data standards as well as proposed and candidate standards to determine if an existing standard element satisfies the data requirement. If no suitable element is found, the developer moves on to the next step.

2. Research of the Element

The purpose of the research is to separate the identity of the data element (what it is) from its applications (how it is used). Research is used to compile adequate information to develop a well-formed domain definition and, if required, a comprehensive list of data values the element may assume. Once a data element's domain has been identified, the appropriate reference element[34] is determined. If no suitable reference element exists, the developer requests a change to an existing reference element or develops and proposes a new one.

3. Definition/Construction of the Element

After the basic research is conducted, the element is documented and submitted for approval. This involves the documentation of the element's attributes, which include its domain, name, qualities, a rigorous definition, and administrative information.

The data element name is developed by applying the Army's naming convention. The data element name format, depicted in Figure 10, is composed of two parts: Prime Term and Reference Element Name. The prime term describes what the data element represents. It is built around a prime word that is the name of an entity in the Army

---

[34]A reference element, also called a "generic element", identifies the structure and physical characteristics of the data values in a domain. Examples of reference elements are Name, Code, and Date.

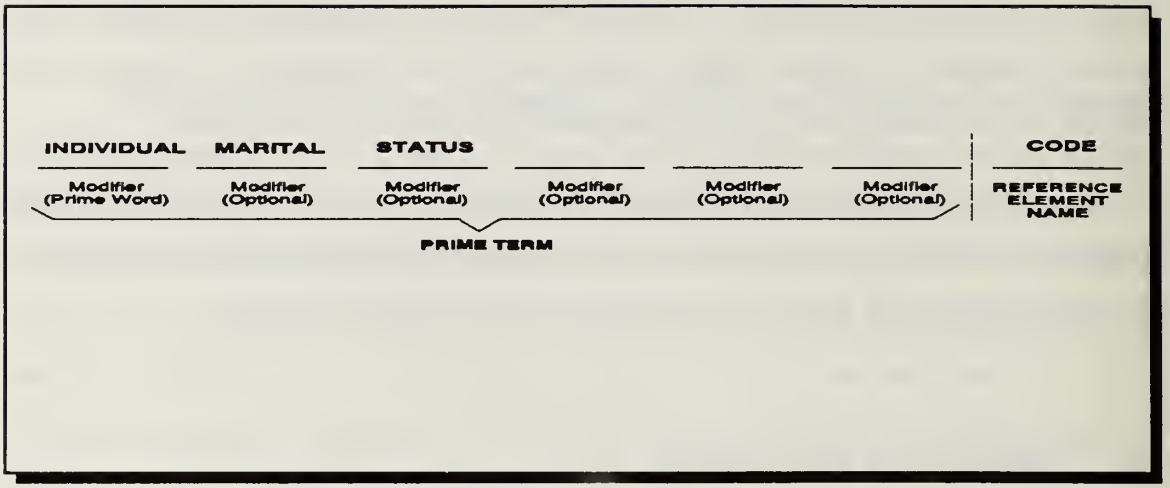

Figure 10. Army Data Element Naming Structure

Data Model. In the example, the data element name *Individual Marital Status Code* is build around the entity *Individual* and describes a particular characteristic of that entity.

The element name must have one word designated as the prime word. The prime word may be in any one of the six modifier positions within the prime term. The data element name can have up to five optional modifiers added to the prime word to fully describe the characteristic the data element represents.

The second part of the data element name is the associated *reference element name*. It identifies the domain values or type of data that can be assigned to the data element. In the example, the reference element name is *Code*. Values assigned to this data element may be S (single), M (married), D (divorced), etc. The data element name must have a prime term followed by a reference element name.

The data element name is the designation given to a data element. This is a descriptive name for reference and is not designed to be incorporated into software applications. To use the data element in a software application, the data element is assigned mnemonic abbreviations. These are unique short forms of the data element name. There are three types of mnemonic abbreviations — short mnemonic abbreviation

(8 or fewer characters), regular mnemonic abbreviation (18 or fewer characters), and long mnemonic abbreviation (32 or fewer characters).

4. Validation of the Element

The developer must ensure that the element can be used universally across all echelons of the Army and DoD. Standardization efforts require coordination with subject matter experts and functional proponents to determine if data requirements have been adequately met. If other data standards are impacted, organizations must work jointly to bring affected data into compliance with data standards. Organizational data administrators will ensure that standard elements are being used.

F. DOA Data Management Tools

The Army's data management program could not have been implemented without the use of a set of tools that would allow them to automate their modeling and standardization methodologies. To assist in the standardization, documentation, and storage of data, the Army developed the Army Data Dictionary/Automated Dictionary Support System (ADD/ADSS). To support their modeling efforts, the Army is in the process of developing the Army Data Encyclopedia. In the interim, the Army has been able to utilize an encyclopedia that was developed by the Army Corps of Engineers. The capabilities of the ADD/ADSS and the future ADE are discussed below.

1. Army Data Dictionary/Automated Dictionary Support System

The ADD/ADSS provides a mechanism to capture and merge the information the Army needs to manage its data resources effectively. The ADD/ADSS is a repository used to approve and record standard elements to be used in Army information systems. Users and systems developers can access the dictionary to find existing, proposed, candidate, or standard elements. New data elements can be submitted on-line or in batch

as candidate or proposed elements. Information Class Proponents and the Army Data Administrator can perform on-line functional and technical approval, respectively. Users of the dictionary can also obtain reports and perform queries.

2. Army Data Encyclopedia

The ADD is actually the preliminary form of a much more sophisticated tool called the Army Data Encyclopedia (ADE), which is currently planned for development by the Information System Engineering Command. The ADE design will be consistent with the IRDS standard and will support the Army objective to field and operate interoperable and integrated systems.

The ADE will support and enhance the effective and efficient management of the creation, documentation, use, modification, maintenance, standardization, and disposition of shared data throughout the Army. Once completed, the ADE will provide an integrated repository of metadata to include architectures, data models, internal models, external views, data elements, directory information activity models, organizational models, and other metadata required.

G. Current DOA Data Management Implementation Status

The Army began implementing their data management program in the Spring of 1990. Since that time, 19 functional areas that represent the way the Army does business have been identified. These functional areas are listed in Table 7. The Army expects to identify additional functional areas as they derive more detailed models. These functional areas will likely be identified by components that receive guidance from the Joint arena (i.e., Special Forces) and therefore were missed during high-level (strategic) modeling.

Modeling teams consist of no more than ten people and include a mix of functional experts, who know the organization's business, and technical experts, who

• Inspector General	• Deputy Chief of Staff for Personnel
• Judge Advocate General	• Director of Management
• Financial Management	• Director Information Systems for Command Control and Communication
• Research and Development Acquisition	• Office of the Director Under Secretary of the Army
• Chaplains	• Program Analysis and Evaluation
• Corps of Engineers	• Secretary of Army Management and System support
• Deputy Chief of Staff for Intelligence	• Surgeon General
• Deputy Chief of Staff for Logistics	• Army Audit Agency
• Deputy Chief of Staff for Operations and Plans	• Chief of Legislative Liaison
	• Chief of Public Affairs

Table 7. Army Functional Areas

understand the modeling methodology and are familiar with modeling techniques. Personnel selected to participate in modeling a particular functional area undergo a one to two-week training course. This course presents an overview of the National Institute of Standards and Technology (NIST) standards for data modeling, introduces the IDEF methodology, and teaches participants how to build functional and data models.

The Army schedules a five-week block of time for a functional area modeling session. With the exception of Logistics, this time frame has provided Army

organizations with enough time to complete strategic modeling and, in some cases, to bring the model down to mid-level management (tactical) or operational activities. The Logistics functional area is so large that their first strategic modeling session resulted in a model that was recognized as too high level to be useful. Logistics was scheduled for an additional 10 week modeling session in order to capture all strategic data and begin mid-level modeling.

Another problem the Army had with some of their initial modeling sessions was that functional area managers did not carefully select appropriate personnel to participate in the modeling session. The result was modeling teams made up of personnel who were experts in only a few activities within their functional area; as a result some activities within the functional area were missed. This problem arose because functional managers did not understand the modeling methodology and/or had low expectations with respect to the benefits of modeling.

The Army has corrected the modeling team staffing problem by selecting personnel who have a broad knowledge of the functional area and supplementing the team with subject area experts (personnel who are experts in one particular activity within the functional area) when required. Most functional managers now have a better understanding of the modeling process and recognize that modeling will assist in identifying areas in which budget cuts can be made.

Currently, 16 of the 19 functional areas have completed at least strategic modeling. The remaining functional areas (Army Audit Agency, Chief of Legislative Liaison, and Chief of Public Affairs) are awaiting funding and are expected to complete strategic modeling by November 1992. Some functional areas have continued with the modeling process and are at various stages of mid-level management or operational modeling.

Management support for process and data modeling depends on a number of factors:

- New information system development requiring more in-depth modeling to identify data needs. For example, certain activities within Operations and Personnel have been brought down to mid-level or operational activity functional

and data models to support the Reserve Component Automation System (RCAS) development.

- The ability of the modeling team to complete the model because of their functional simplicicty (Chaplains).

- Functional experts desire to complete modeling because they have mastered the modeling techniques. This is happening with the Logistics functional area.

- Functional proponents belief that functional and data modeling is extremely valuable in identifying areas that they can apply re-engineering techniques in the wake of budget decreases.

The Army has integrated 14 of the 16 functional area data models into the Army Data Model. Because personnel responsible for integrating functional data models into the Army Data Model are also assisting functional areas in their modeling effort, integration lags functional modeling by about a month.

The modeling effort has thus far identified over 650 entities, of which 340 are approved. There are currently more than 2450 data elements in the ADD approximately 30% have been standardized and the remaining 70% are still undergoing evaluation. The Army estimates that on average, every data element identified and standardized could replace eight synonym data elements that exist in current information systems.

The Army's goal is to standardize an entity within 30 days and a data element within two weeks, including functional review and approval by the Information Class Proponent (ICP) and technical review and approval by the Army data administrator. Initially, data entities required five to six months and elements up to four months to standardize. Now that ICP's and the Army data administrator staff have a better understanding of their roles, know how to use the ADD/ADSS, and have mastered the Army's naming convention, the Army has reduced the entity standardization to four months and data element standardization time to 30 days. There are three main reasons why the Army has not yet met their standardization time goals:

- Strategic modeling identifies a large portion of the entities that are of importance to the organization. Therefore, the ICP's and data administrator staff have been inundated with a large number of entities and key elements in a short period of time.

- Data element identification and naming hinges on entity approval, and so the Army is taking time to ensure that data entities are well formulated.

- Army data administrator staff personnel are currently spread thin trying to support the modeling development and integration effort, conduct Army to DoD data model comparisons, and perform technical review and approval of data entities and elements.

The Army expects to complete at least mid-level modeling by September 1994. This target date depends on the availability of personnel and budgetary resources. Operational level modeling will be prioritized for functional area activities as required by new information system development or as requested by functional area managers if budgetary resources are available. The Army Information Architecture is also evolving. The Information Architecture's information model and functional and data architectures reflect most of the strategic modeling effort that has been completed thus far. The Geographic/Technical Architecture is on hold until the Army can identify a tool that will assist in its development.

It is important to realize that completion of the modeling process and the integration of those models into the Army Data Model and Information Architecture provides the Army with a baseline representation of their functional processes and the data that are created and used by functional areas. As re-engineering techniques are applied and new information systems are developed, the functional models, and to a lesser extent the data models, will have to be updated to reflect those changes. Thus, the Army's initial modeling effort is only the beginning of a continual process. Long-term benefits of this process can be realized only if the models and architectures are kept up-to-date.

V. Lessons Learned

A. Commitment by Management is Essential

Management commitment is essential when organizations undergo change. This commitment must be shared by upper management as well as functional managers. Upper management commitment ensures that overall guidance is provided and that resources will be available to institute the change. Functional management commitment ensures the change is implemented successfully.

Gaining commitment from management has not been easy for the Army. A 1990 Government Accounting Office (GAO) report stated that the Army's efforts to improve management and acquisition of its information resources had not been fully successful.[30] According to the GAO, the Army did not adequately pursue the development of an information architecture and cited a lack of local commanders' commitment as one of the problems.

Although there still remain pockets of contention, the majority of Army management has found that their IRM and data management program efforts will allow them to continue their mission in an environment where downsizing and shrinking budgets are a continuing fact of life.

B. Data Manager Must be in a Position of Authority

In order for a data management program to be successful, the data manager must be in a position of authority to set policy and ensure the appropriate plans and controls

[30]GOA/IMTEC 90-58

are implemented. All too often the data manager is placed several levels down in the organization and has little input in strategic decision making. The Army has ensured that data management issues are addressed by giving DISC4 overall responsibility for the Army's IRM and data management programs.

C. Effective Communication is Key

When the policy and implementation functions of data management are separated (both geographically and functionally), there must be a close working relationship between those who set policy and those who implement it. As with any military organization, Army policy is set by upper management and implemented by functional managers much lower in the chain of command. The Army has found that without effective communication, a well intentioned policy will be difficult to implement in the intended manner. A majority of Army managers have stated that they wish they understood the value of the Army information model five years ago. As one Army officer put it "We had a vision, we just did not do a good job of articulating it."

D. Technical Tools are Invaluable

The data manager, data administrators, and database administrators must have the technical tools available to implement the data management policy. Tools such as the data dictionary, data encyclopedia, and I-CASE environments provide the means to automate data management methodologies as well as assist in the development and maintenance of strategic information systems.

The Army saw the need for such tools long before they began implementing their data management program. The Army Data Dictionary/Automated Dictionary Support System has been an invaluable tool for the Army not only for storing standard entities and elements and their attributes, but also for automating the approval process. Although the Army Data Encyclopedia is not complete, the Army has been able to use the Army

Corps of Engineers encyclopedia as a mechanism to define and store their process and data models under IDEF modeling techniques. The ADD/ADSS and the Corps of Engineers encyclopedia are stand-alone tools. Currently the Army does not have the capability to link data entities and elements in the dictionary with the models contained in the encyclopedia. The ADE planned should provide this integration in the future.

E. Model the Data before Standardizing

An organization needs to perform process modeling in order to understand what business it is in and what data it uses. By modeling, the organization can determine what data are essential to the organization and who uses them. Data standardization and naming conventions use the information obtained from modeling to properly define elements, name them, identify their attributes, and assign lifecycle responsibilities.

Initially, the Army Data Management and Standards Program stated that data element standardization begins when data elements required to support applications are identified during development of an information system. The Army began standardizing personnel elements needed for the Standardized Installation/Division Personnel System (SIDPERS) using AR 25-9 procedures.

The Army soon realized that standardizing entities and elements for SIDPERS would not help in identifying what other functional areas required access to portions of the personnel data or if data values should be captured by another functional area and shared with personnel. This problem occurred because the data management program focused on data elements: it provided policies for naming data elements, but did not address procedures for identifying strategic data. The Army data management program has been revised to include data modeling as a basis for the identification of Army data.

F. End-Users Play a Critical Role

End-users must be involved with data management activities. They are the resident experts on how the business operates and what data are required to perform the job. They can provide valuable insight for the data modeling process and directly affect the success of any data management program. The Army's modeling teams directly involve the end-user. However, the Army found that selecting the right end-users to participate is just as important.

Initially, functional area managers did not carefully select personnel to participate in the modeling process. This practice resulted in incomplete models because modeling team members could not lend expertise in all activities within the functional area. Selecting the wrong personnel was caused by a lack of understanding by management regarding the modeling process, as well as low expectations by management with respect to overall benefits.

Fortunately, the Army identified this weakness after initial modeling activities. The Army now stresses the use of functional experts who possess a broad understanding of the organization and its mission, as well as an overall functional knowledge. Subject matter experts are called in as necessary to support the modeling effort and to review modeling products. By ensuring that end-users, and in particular, experts, play an active role in the data management effort, the Army has paved the way for a successful program.

G. Data Management Programs Take Time and Resources

An organization cannot be expected to revolutionize the way they handle their data resources overnight. Many organizations hesitate to invest resources for a program that does not yield immediate benefits. A sound data management program provides long-term benefits, but requires an up-front contribution of personnel, time, and budgetary resources.

The Army Data Management Program began in 1983 when the Army realized that conflicting and inconsistent data were having an adverse effect on Army decision making.

It has taken nine years for the Army to develop policies, procedures, and tools that will allow it to manage its data resources effectively.

The Army has not completed initial implementation of its data management program (i.e., modeling the 19 functional areas). The reasons for this are numerous:

- Budgetary constraints
- Personnel constraints
- Lack of initial commitment to the program
- Lack of initial expertise in data management methodologies and modeling
- Size and scope of the DOA

However, the Army has made great strides and have already received some benefits from their data management program by identifying redundant activities within functional areas which may be eliminated.

H. Data Management and IRM is a Continual Process

Developing an initial set of models and architectures and standardizing the data identified during modeling is only the beginning of the data management and IRM process. Organizations are not static; they change as the environment in which they operate changes. As organizations take advantage of the information obtained from the modeling effort and employ re-engineering techniques, or develop new information systems and apply new technology based on the plans produced from the IS architecture, the functional and data models and IS architecture must be updated to continue to benefit the organization.

The Army is just beginning this process. Long-term benefits are still on the horizon, but are achievable if the Army continues to apply and enforce their data management and IRM strategies. For other DoD organizations that have not yet begun developing or implementing a comprehensive data management and IRM program, they

can shorten the process by learning from the Army's mistakes and repeating their successes.

Glossary of Terms

ADD/ADSS - Army Data Dictionary/Automated Dictionary Support System.

ADE - Army Data Encyclopedia.

APPLICATIONS ARCHITECTURE - A framework that represents computer programs that perform tasks associated to a particular business function.

ARCHITECTURE - A framework for specifying how components of a system fit together.

CASE - Computer-Aided Software Engineering.

CIM - Corporate Information Management.

COMMUNICATIONS ARCHITECTURE - A framework for developing a communication network to link hardware, software and information.

DATA - Meaningful facts about persons, places, things, concepts, events, and activities in a defined format and structure from which information may be derived.

DATABASE ADMINISTRATOR - The organization's technical expert on database related activities, who is also responsible for the operation of the databases.

DATA ELEMENT - A characteristic or attribute of an entity. The lowest level of addressable data.

DBMS (DATABASE MANAGEMENT SYSTEM) - A set of programs that are used to define, process, and administer the data base and its applications.

DATA ADMINISTRATOR - A person who has overall responsibility for the organization's data resources.

DATA ARCHITECTURE - A framework for organizing and defining the interrelationships of data in support of an organization's information architecture. The data that will be used to bind all the other architectures together.

DATA DICTIONARY - A centralized repository of information about data.

DATA DICTIONARY SUBSYSTEM - Facilitates storage of the organization's metadata and has the capability to relate that information in such a way that the organized metadata serves a useful purpose.

DATA MANAGEMENT - The process of locating, organizing, cataloging, storing, retrieving, and maintaining data which is fundamental to the organization.

DATA MODEL - Describes the entities, their attributes and interrelations that are necessary to support the business processes.

DATA STANDARDIZATION - Provides a common structure which involves naming the entity using the organization's naming convention, providing a precise definition, identifying its characteristics (attributes), and developing access keys.

DISC4 - Director Information Systems for Command, Control, Communications and Computers.

DOA - Department of the Army.

DoD - Department of Defense.

DOMAIN - Specifies format, length, and special restrictions on the value of each domain.

ENCYCLOPEDIA - An extension of the data dictionary used to develop and store data models and architectures and documents their relationship to the data stored in the data dictionary.

END-USER - A collective term used for anyone who uses data and applications to provide information.

ENTITY - Any person, place or thing that has meaning to a user.

ENTITY-RELATIONSHIP DIAGRAM - A diagram which depicts the relationships between entities (ie. not related, one-to-many, one-to-one).

FILE PROCESSING SYSTEM - A computer system which has corporate information embedded within the application program's source code.

FUNCTIONAL AREA - Any area within an organization that has a definable set of tasks.

FUNCTIONAL MODEL - Depicts the activities or processes an organization performs to accomplish its mission.

HARDWARE/SOFTWARE ARCHITECTURE - A framework to provide the processing power needed to run applications that will generate and distribute information.

INFORMATION ENGINEERING - A process which analyzes the strategic information needs of the organization.

INFORMATION SYSTEM - Activities and resources concerned with the creation, gathering, manipulation, classification, storage and transmission of elements of information.

INFORMATION SYSTEMS ARCHITECTURE - A framework within which future IS development, procurement, and implementation activities can occur.

INFORMATION RESOURCE MANAGEMENT (IRM) - The process of directing or controlling the use of an information system comprised of any combination of hardware, software, procedures, documents or people that transforms data into a meaningful and useful form for satisfying organizational goals and objectives.

METADATA - Data about the structure of data stored in the data dictionary. It includes the data name, format, domain as well as other characteristics.

SCHEMA - The structure of the database or how data is physically stored.

STOVEPIPE SYSTEM - A system developed to meet one functional area without taking into account other functional areas which potentially use the same information.

STRATEGIC DATA PLANNING - A top-down strategy for identifying and understanding data in the context of the overall business functions.

STRATEGIC ENVIRONMENT - Within information mission area (IMA), the environment which links the Theater/Tactical and Sustaining Base Environments.

SUBSCHEMA - A family of applications to process portions of the data or how the user views the data.

SUSTAINING BASE ENVIRONMENT - The area and information resources outside of the area of operations that have the responsibility to raise, organize, train, equip, and eventually, deploy and sustain Army forces in the accomplishment of their mission in operational theaters.

THEATER/TACTICAL ENVIRONMENT - Army area of operations.

USAISC - United States Army Information Systems Command.

References

Alavi, Maryam, et al., "Strategies for End-User Computing: An integrative Framework", *Journal of Management Information Systems*, 4, 3, Winter 1987-88.

Aranow, Eric, "Full Lifecycle Case: Views from the Tower of Babel", *System Builder*, October/November 1989.

Brewin, Bob, "Federal Computer Week White Paper: CIM", *Federal Computer Week*, September 1991.

Brown, C. and Bostrom, R., "Effective Management of End-User Computing: A Total Organization Perspective", *Journal of Management Information Systems*, 6, 2, Fall 1989.

Champy, James A., "Mission Critical", *CIO*, January 1992.

"Corporate Information Management Process Improvement Methodology for DoD Functional Managers", D. Appleton Company - Unpublished.

Finkelstein, Clive, "Together at Last", *Computerworld*, 16 December 1991.

Gacel, Rafael A., *The Top Ten Critical MIS Issues in the Department of Defense*, Masters Thesis, Naval Postgraduate School, March 1991.

Gibson, C. F. and Nolan, R. L., "Managing the Four Stages of EDP Growth", *Harvard Business Review*, January/February 1974.

Goodhue, Dale, et al., "Managing the Data Resource: A Contingency Perspective", *MIS Quarterly*, September 1988.

Goodhue, Dale, et al., "Strategic Data Planning: Lessons from the Field", *MIS Quarterly*, 16, 1, March 1992.

Government Accounting Office, "Army Should Limit New Initiatives Until Management Program is Implemented", *GAO/IMTEC-90-58*, June 1990.

Hammer, Michael, "Reengineering Work: Don't Automate, Obliterate", *Harvard Business Review*, July/August 1990.

Head, Robert V., "Looking Back at Two Acts that Shaped Federal IRM", *Government Computer News*, 9, 4, 12 November 1990.

Headquarters Department of the Army, *Army Data Management and Standards Program (AR 25-9)*, 25 September 1988.

Headquarters Department of the Army, *Army Data Management and Standards Program: Administrative Procedures Guide (DA Pam 25-9-1)*, 4 June 1992.

Headquarters Department of the Army, *Army Information Architecture (DA Pam 25-1)*, 20 August 1991.

Headquarters Department of the Army, *The Army Information Resources Management Program (AR 25-1)*, 18 November 1988.

Headquarters U.S. Army Information Systems Command, "Program Plan: Army Information Engineering, Standard Army Management Information Systems Modernization, Army Executive Software, Reusable ADA Products for Information Systems Development, and The Army Data Management and Standards Program", Volume 1, February 1991.

Kanter, Jerry and Miserendino, Joe, "Systems Architectures Link Business Goals and IS Strategies", *Data Management*, November 1987.

Kroenke, David M. and Dolan, Kathleen A., *Database Processing: Fundamentals, Design, Implementation*, Science Research Associates, Inc., 1988.

Narayan, Rom, *Data Dictionary: Implementation, Use, and Maintenance*, Prentice Hall, Inc., Englewood Cliffs, New Jersey, 1988.

Nolan, Richard L., "Managing the Computer Resource: A Stage Hypothesis", *Communication of the ACM*, 16, 7, 1973.

Nolan, R. L., *Management Accounting and Control of Data Processing*, National Association of Accountants, New York, 1977.

Nolan, R. L., "Managing the Crisis in Data Processing", *Harvard Business Review*, March/April 1979.

Rogers, Bill, "SBIS Flattens the Army Hierarchy", *Government Computer News*, 23 December 1991.

Schwartz, Karen D., "Vendors Scramble to Comply With IRDS Standards", *Government Computer News*, 10, 8, April 1991.

Bibliography

Allen, Brandt and Boynton, Andrew, "Information Architecture: In Search of Efficient Flexibility", *MIS Quarterly*, 15, 4, December 1991.

Cage, Glen, "IS Architecture Artistry", *Computerworld*, 29 July 1991.

Emery, James C., *Management Information Systems*, Oxford University Press Inc., New York, New York, 1987.

Finkelstein, Clive, *An Introduction to Information Engineering*, Addison-Wesley Publishing Company, New York, 1989.

Goldfine, Alan, "Proceedings of the Federal Information Processing Standards (FIPS) Workshop on Information Resource Dictionary System (IRDS) Applications", U.S. Department of Commerce, Gaithersburg, Maryland, December 1988.

Kelleher, Joanne, "Data Bases Why They're Finally Usable", *Personal Computing*, 10, 10, October 1986.

Lederer, Albert L. and Sethi, Vijay, "The Implementation of Strategic Information Systems Planning Methodologies", *MIS Quarterly*, September 1988.

Meltzer, Herbert, "An Overview of the Administration of Data Bases", *Database Management Systems*, AFIPS Press, Montvale, New Jersey, 1976.

Distribution List

Agency	No. of copies
Defense Technical Information Center Cameron Station Alexandria, VA 22314	2
Dudley Knox Library, Code 0142 Naval Postgraduate School Monterey, CA 93943	2
Office of Research Administration Code 08 Naval Postgraduate School Monterey, CA 93943	1
Library, Center for Naval Analyses 4401 Ford Avenue Alexandria, VA 22302-0268	1
Department of Administrative Sciences Library Code AS Naval Postgraduate School Monterey, CA 93943	1
Director of Information OASI (C3I) Washington, D.C. 20301-3040	10
Tung X. Bui Code AS/Bd Naval Postgraduate School Monterey, CA 93943	10
Daniel R. Dolk Code AS/Dk Naval Postgraduate School Monterey, CA 93943	10